



Идентификаторы : преодоление языковых барьеров

Ян Писански &

Мая Жумер

Университет Любляны, Словения

Тронд Аальберг

Норвежский университет науки и технологии

Трондхейм, Норвегия

Перевод:

А. Хлебаевой, РНБ

Meeting:

93. Cataloguing

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY
10-15 August 2010, Gothenburg, Sweden
<http://www.ifla.org/en/ifla76>

Резюме:

Идентификация библиографических объектов (entities) даёт возможность беспрепятственно и многократно использовать библиографические данные в различных приложениях, и поэтому она является одной из важных составляющих преодоления лингвистических и культурных барьеров, имеющих в библиотечном мире. В докладе рассматриваются уже существующие потенциальные «кандидаты на роль» идентификаторов формальных объектов (entities) и обсуждается приемлемость каждого из них – с точки зрения широты распространения в библиотеках и за их пределами.

Введение

Т.к. библиотеки имеют дело с нарастающим объёмом информации, важно, чтобы пользователи отчётливо понимали, с какой средой они взаимодействуют. С точки зрения Функциональных требований к библиографическим записям (FRBR) – новой парадигмы, связанной с библиографической вселенной, пользователям необходимо иметь возможность различать примеры (образцы) объектов, чтобы решать и свои пользовательские задачи: находить, идентифицировать, выбирать (отбирать), и любые возможные задачи, не получившие определения в FRBR, - такие как исследования или разыскания.

Необходимость уникальной идентификации частных случаев всех важных сущностей библиографической вселенной постоянно возрастает. Т.к. FRBR - это единственная формально признанная модель библиографической вселенной, кажется естественным, что тем объектам, которые формируют костяк FRBR, следует иметь идентификаторы.

Конечно, FRBR – это модель, которая подаёт себя в различных интерпретациях, и некоторые её дефиниции расплывчаты. Однако это не означает, что не следует делать усилий, направленных на идентификацию различных сущностей. Пусть безупречная идентификация возможна не всегда, - это связано с тем, что природа библиографической вселенной по определению состоит не только из черного и белого цветов, - наличие вместо неё системы определённых идентификаторов сущностей должно помочь пользователям найти собственный путь, вне зависимости оттого, как кто-то интерпретирует библиографическую вселенную.

Необходимость идентификации в библиотеках

Библиотеки традиционно полагались на сочетание атрибутов, как на единственную возможность идентифицировать библиотечные материалы и другие объекты (например, рубрики для имён лиц, унифицированные рубрики). Хотя в пределах традиционной библиотечной среды и в ограниченной среде это работает относительно хорошо, библиотеки больше не могут позволять себе избыточно расходовать силы, дублируя каталогизацию и авторитетный контроль, если эти библиотеки изолированы друг от друга.

Помимо того, во многих случаях применение идентифицирующих элементов в отдельных каталогах было неудачным, результатом чего становились непоследовательно представленные данные. Стоимость труда относительно высока, а распространение использования информации и коммуникационные технологии предоставляют средства для эффективного и целесообразного разделения труда даже на всемирном уровне. Пусть местные идентификаторы обычно достаточно хороши для местного применения, - истинная сила идентификации заключается в идентификаторах, признанных во всём мире. Пользователи не желают ограничивать свой поиск одной библиографической базой данных и всё чаще требуют интегрированного доступа к широкому ряду он-лайнных ресурсов, таким образом, вынося на рассмотрение вопрос о гармонизации библиографической информации. И, наконец, библиотекам необходима возможность интегрировать в возникающие социальные сети, стать их частью и получить возможность собирать и распространять поставленный пользователями контент, - как дополнительный ресурс.

Однако, при таких обстоятельствах идентификаторы, зависящие от культуры (включая идентификаторы, зависящие от языка и графики), связаны присущими им недостатками. Предполагается, что пользователи хотят, чтобы при выводе информации применялись решения, приемлемые для данной культуры, - и так оно и должно быть. Однако необходимость выводить информацию на показ так, чтобы её было легко идентифицировать, и необходимость уникальным образом идентифицировать частные случаи сущностей – это два разных вопроса. Применение имен (наименований) для международной идентификации любого вида объекта – не лучшее решение. Дело не только в том, что имена (наименования) могут меняться и это обстоятельство ставит под сомнение их применение для уникальной идентификации; кроме того, существует проблема имён (наименований), зависящих от культуры и языка.

Возможно, некоторые считают, что ради идентификации все могли бы применять одну и ту же форму имени (наименования). Такое решение ведёт к тому, что одна культура подавляет все остальные. Непросто будет найти решение, приемлемое для всех. Фактически, нет никакой необходимости в идентификации, проводимой подобным способом. Рабочая группа Библиотеки Конгресса «Будущее библиографического контроля» сформулировала это так: «Применение языковых цепочек [...] в качестве

идентификаторов и для показа, и для применения данных, препятствует обмену данными между языковыми сообществами и другими сообществами, применяющими различные данные» (2008, стр. 24)

Идентификаторы могут стать базисом авторитетных файлов, и таким образом могут помочь сократить избыточность в каталогах, а также – значительно упростить использование каталогов. Действительно ли авторитетные файлы помогают пользователям? – Ответ зависит в значительной степени от применения в каждом частном случае. Есть случаи, когда авторитетные записи не связаны с библиографическими. Решение, при котором изменения имени автора (например, в случае вступления в брак), ведёт к необходимости вручную изменять все эти данные в библиографических записях, не может быть приемлемым.

Кроме того, идентификаторы становятся ещё важнее, при зарождении новых способов применения и многократного использования данных, поставляемых библиотеками.

Например, идентификаторы являются ключом к успешной интеграции библиографических данных в Semantic Web (семантический веб) и заинтересованность библиотек в использовании Linked Data постоянно возрастает.

Хотя потребность библиотек в идентификации не является новостью (Tillett, 2007), в последнее время электронный век вызвал к жизни множество разных инициатив, некоторые из которых уже описаны: Vitiello (2004), Nakala (2006), Tillett (2007) и Babeu (2008).

Идентификаторы

Идентификация объектов – одна из наиболее интригующих областей применения идентификаторов в библиотеках и за их пределами. То, что Рабочая группа ИФЛА по Функциональным требованиям к авторитетным записям и их нумерации занималась оценкой возможности введения Международного стандартного номера авторитетных данных (ISADN), – яркое свидетельство признания библиотечными кругами значения правильной идентификации, выходящей далеко за пределы объектов Группы 2.

Однако, группа пришла к выводу, что введение такого номера невозможно, хотя в самом отчёте (Tillett, 2008) не содержится подробных экономических обоснований, – только вывод.

В этом отчёте обсуждались три возможности. Помимо отвергнутого стандартного номера ISADN, признанного, в общем, неплохим, но дорогостоящим, с полным основанием отвергнуто решение, выполнение которого трудно постоянно поддерживать: идентификация на основе текста (единственный принятый заголовок). Закладка в основу идентификации кластеров – авторитетных файлов, полученных из разных источников (или проект Виртуального авторитетного файла) в результате обсуждения была оценена как хороший вариант для будущих разработок. Отчёт также признает продолжающуюся разработку ISO - ISNI (Международный стандартный идентификатор имён International Standard Name Identifier) в качестве стандартного идентификатора, который может быть полезен некоторым сообществам.

Важно узнать, могут ли существующие идентификаторы шире использоваться в библиотеках, хотя они просто несправедливо забыты, нужны ли новые идентификаторы в свете FRBR и их применения в библиотеках. Последовательное применение идентификаторов, независимо от их формы, но предпочтительно – в высшей степени стандартизированных, должно стать краеугольным камнем не вполне ясного будущего – параллельного использования «унаследованных» данных и данных,

порожденных FRBR. Идентификаторы должны также открыть новые пути многократного использования библиографической информации и связи библиографической информации с другими ресурсами, упрощения взаимодействия и совместимости с другими доменами. Не следует также забывать, что целесообразности широко используемых идентификаторов могут препятствовать культурные различия, которые дают себя знать при интерпретации библиографической вселенной.

Уже существует несколько международных идентификаторов, нацеленных на идентификацию библиографических материалов. Однако внимание их авторов было сосредоточено не на объектах FRBR, а на попытках удовлетворить разные потребности разных сообществ. Из-за этого возможны трудности при безусловном определении того, какие именно объекты FRBR они пытаются идентифицировать (как можно увидеть в Таблице 1) - оценки каких свойств даны несколькими известными авторами. Противоречия в этой таблице можно объяснить тем, что оценка была дана до официального признания некоторых из этих стандартов. Итак, как явствует из таблицы, наименее проблематичными являются идентификаторы воплощения.

Таблица 1: Сравнение оценки идентификации объектов 1-ой группы FRBR (W=произведение, E = выражение, M = воплощение)

	VITIELLO (2004)	GATENBY (2008)	LEBOEUF (2005)	HAKALA (2006)
ISBN	M	M	M	M
ISSN	M	M	M	M
ISRC	E	M	E	
ISAN	W, E*	W	W	
ISWC	W	W	W, E	
ISTC		W	E	W, E
ISMN	M	M	M	
V-ISAN		M	E	

*Возможно, автор фактически ссылается на ISAN, идентифицирующий произведения, и V-ISAN, идентифицирующий выражения, хотя из текста это неясно.

Идентификаторы проявления до сих пор не являются наиболее широко применяемыми в библиотеках идентификаторами объектов. Поэтому грустно осознавать, что только 30% материалов WorldCat имеют международный идентификатор (Gatenby, 2008). Если считать, что данные WorldCat правильно представляют состояние современных библиографических данных в мире, этот показатель должен вызвать большую озабоченность, т.к. он указывает на отсутствие готовности, даже когда дело касается проявлений.

Итак, относительный успех ISBN и, в меньшей степени, - других идентификаторов проявления (таких как ISMN), показывает, что можно проводить идентификацию на всемирном уровне, если внедрять её правильно. Её принятию тихо способствовали все, кто имеет дело с книгами, а обсуждения стоимости системы, которая обеспечивает понятную идентификацию, возникают редко. Но по причине охвата (например, материал, возраст и географическое расположение) некоторое количество воплощений остается без ISBN, т.е. это решение определённо имеет недостатки. Holdsworth (2008) указывает ещё одну проблему: в условиях FRBR одни и те же номера ISBN иногда присваиваются различным воплощениям одного и того же выражения.

Если идентификация воплощений относительно понятна, идентификация произведений и проявлений не так ясна, особенно из-за того, что библиографическая вселенная охватывает все типы материалов. Кроме того, на этих - более абстрактных - уровнях идентификаторы редко фигурируют в самих воплощениях. К сожалению библиотек, слишком часто информация, которой нет в самом проявлении, не существует вообще, даже если она представляет большой интерес для пользователя. Не следует также забывать, что до сих пор вся каталогизация производилась и происходит на уровне воплощения, и необходимость идентифицировать произведения и выражения была менее актуальной. Из-за всего этого в настоящее время идентификации произведений и выражений в библиотеках почти не существует. ISRC (International Standard Recording Code = Международный стандартный код записи) является идентификатором выражения, а ISAN (International Standard Audiovisual Number = Международный стандартный аудиовизуальный номер) является идентификатором произведения, однако они связаны со сравнительно малой частью библиографической вселенной и в библиотеках используются не особенно хорошо. С другой стороны, сущность ISMWC (International Standard Music Work Code = Международного стандартного кода музыкального произведения) и ISTC (International Standard Text Code = Международного стандартного кода текста) не настолько очевидна. Согласно организации ISWC, собственные ISWC получают адаптации и переводы (Antelman, 2004). Фактически, база данных ISWC (www.iswc.org) демонстрирует в высшей степени непоследовательное присвоение кодов ISWC, с точки зрения уровней FRBR, а также – с точки зрения географического охвата.

С другой стороны, стандарт ISTC демонстрирует внутреннюю непоследовательность. Раз уж ISTC применим к “текстовым произведениям”, а тексты в FRBR ассоциируются с выражениями, коды ISTC должны присваиваться на этом уровне. Фактически, примеры, приводимые в Приложении Е стандарта ISTC (ISO 21047, 2009), являются различными вариантами одного и того же текстового произведения (например, новые редакции и переводы - пример Е.4), которым присвоены разные номера ISTC. Отсюда следует, что фактически номера ISTC применяются с точки зрения FRBR на уровне выражения, а не произведения. Но это прямо противоречит положению Приложения В (В.13) опубликованного стандарта ISTC: одному и тому же текстовому произведению не может быть присвоено больше одного ISTC. Хотя в тексте примера Е.4 описаны связи между одним текстовым произведением и несколькими текстовыми произведениями, произошедшими от первого, приводимые примеры ясно указывают, что переводы и редакции по-прежнему рассматриваются, как то же текстовое произведение, а адаптация для детей «оригинала» текстового произведения – нет.

Т.к. мы живём в электронном мире, где идентификация электронных и не-электронных объектов одинаково важна, мы не должны забывать о таких идентификаторах, как DOI. DOI (Digital Object Identifier = Идентификатор электронного объекта) – это относительно хорошо используемый электронный идентификатор объектов интеллектуальной собственности. Это, однако, не предусматривает идентификации каждой отдельной сущности. Фактически, можно идентифицировать физические или электронные воплощения, представления и абстрактные произведения (International DOI Foundation, 2006). Если ориентироваться на точку зрения FRBR, это означает, что даже хорошо задуманное использование DOI в действительности может привести к большим недоразумениям.

Заключение

Имеется несколько различных международных стандартов для идентификации различных частей существующей библиографической вселенной. И хотя некоторые применяются хорошо и применимы к частному случаю сущности FRBR (например, ISBN, ISMN), большая часть зацепок для идентификации сущностей FRBR на международном уровне – совершенно не на месте. Ещё важнее то, что существующие идентификаторы предназначены для идентификации только фрагментов библиографической вселенной (например, текст, музыка, АВМ), используются мало, часто даже неправильно (возможно, из-за культурных различий) и иногда даже используются для того, чтобы идентифицировать *частные случаи* различных сущностей FRBR, не обладая механизмом для должной идентификации *самих* сущностей. К тому же, в некоторых случаях оказывается, что библиотекам трудно или слишком дорого получить доступ к существующим идентификаторам. Всё это заставляет сомневаться в том, что в ближайшем будущем возможна эффективная идентификация сущностей FRBR с использованием существующих идентификаторов. С другой стороны, аналогичные проблемы, похоже, должны возникнуть при принятии любых новых идентификаторов.

Библиотечное сообщество уже доросло до того, чтобы принять идентификаторы, которые могут совместно использоваться разными сообществами во всем мире, что является жизненно важным для всемирной интеграции библиотек. Остаётся только смотреть: что это будут за идентификаторы, что они будут идентифицировать. Пока затраты, связанные с поддержанием системы идентификации, способной оказывать подобные услуги, высоки, а точная идентификация не всегда возможна, существующее положение дел может привести к ещё большим затратам.

Библиография

Antelman, K. (2004). Identifying the Serial Work as a Bibliographic Entity. *Library Resources & Technical Services*. 48 (4), 238-255.

Babeu, A. (2008). *Building a »FRBR-Inspired« Catalog: The Perseus Digital Library Experience*.

<http://www.perseus.tufts.edu/~ababeu/PerseusFRBRExperiment.pdf>

Gatenby, J. (2008). *The activities of OCLC on FRBR*. Workshop on FRBR in the European Library, 9. October 2008, Lisbon, Portugal.

http://frbr.bnportugal.pt/documentos/The_activities_of_OCLC_on_FRBR.ppt.

Hakala, J. (2006). The seven levels of identification. *Program*. 40 (4), 361-371.

Holdsworth, M. (2008). *The Identification of Digital Book Content*. Report prepared for the Book Industry Study Group, January 2008.

http://www.bisg.org/docs/DigitalIdentifiers_07Jan08.pdf

International DOI Foundation (2006). *DOI Handbook*.

<http://www.doi.org/hb.html>

ISO 21047 (2009). *Information and documentation – International Standard Text Code (ISTC)*, 22 p.

LeBoeuf, P. (2005). *Identifying 'textual works'*. FRBR in 21st Century Catalogues, Dublin, Ohio, May 2-4 2005.

<http://www.oclc.org/research/events/frbr-workshop/presentations/leboeuf/ISTC.ppt>

Library of Congress (2008). *On the record: Report of The Library of Congress Working Group on the Future of Bibliographic Control*.

<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>

Tillett, B. (2007). Numbers to Identify Entities (ISADNs – International Standard Authority Data Numbers). *Cataloging and Classification Quarterly*. 44 (3/4), 343-361.

Tillett, B. (2008). *A Review of the Feasibility of an International Standard Data Authority Number (ISADN)*. Prepared for the IFLA Working Group on Functional Requirements and Numbering of Authority Records, edited by G. Patton.

<http://archive.ifla.org/VII/d4/franar-numbering-paper.pdf>

Vitiello, G. (2004). Identifiers and identification systems. *D-Lib Magazine*. 10 (1).

<http://www.dlib.org/dlib/january04/vitiello/01vitiello.html>

International Standards Discussed in the Paper

ISO 2108 (2005). *Information and documentation – International Standard Book Number (ISBN)*, 21 p.

ISO 3297 (2007). *Information and documentation – International Standard Serial Number (ISSN)*, 20 p.

ISO 3901 (2001). *Information and documentation – International Standard Recording Code (ISRC)*, 9 p.

ISO 10957 (2009). *Information and documentation – International Standard Music Number (ISMN)*, 13 p.

ISO 15706-1 (2002). *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 1: Audiovisual work identifier*, 12 p.

ISO 15706-2 (2007). *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 2: Version Identifier*, 20 p.

ISO 15707 (2001). *Information and documentation – International Standard Music Work Code (ISWC)*, 10 p.

ISO 21047 (2009). *Information and documentation – International Standard Text Code (ISTC)*, 22 p.